

Arithmetic Progressions with Restricted Digits

Aled Walker and Alexander Walker

Abstract. For an integer $b \geq 2$ and a set $S \subset \{0, \dots, b-1\}$, we define the *Kempner set* $\mathcal{K}(S, b)$ to be the set of all nonnegative integers whose base- b digital expansions contain only digits from S . These well-studied sparse sets provide a rich setting for additive number theory, and in this article we study various questions relating to the appearance of arithmetic progressions in these sets. In particular, for all b we determine exactly the maximal length of an arithmetic progression that omits a base- b digit.

1. INTRODUCTION. In 1914 Kempner [6] introduced a variant of the harmonic series that excluded from its sum all those positive integers that contain the digit 9 in their base-10 expansions. Unlike the familiar harmonic series, Kempner's modified series converges (the limit later shown to be approximately 22.92, see [1]). A simple generalization of Kempner's original argument shows that convergence occurs as long as any nonempty set of digits is excluded, and that this result holds in any base (see [10], for example).

Let us introduce some notation to describe these results in general. Fix an integer $b \geq 2$ and a subset of integers $S \subseteq [0, b-1]$. Here and throughout the article, for two integers x and y we use $[x, y]$ to denote the set $\{n \in \mathbb{Z} : x \leq n \leq y\}$. We then define the *Kempner set* $\mathcal{K}(S, b)$ to be the set of nonnegative integers that, when written in base b , contain only digits from S . Thus $\mathcal{K}([0, 8], 10)$ denotes the set originally studied by Kempner. We will assume throughout that $0 \in S$, to avoid the ambiguity of leading zeros, and require $S \neq [0, b-1]$, to preclude the trivial set $\mathcal{K}([0, b-1], b)$ (which is nothing more than $\mathbb{Z}_{\geq 0}$). These sets S will be referred to as the *permitted sets* S , and the related Kempner sets $\mathcal{K}(S, b)$ as *proper Kempner sets*.

The arithmetic properties of proper Kempner sets have been the object of considerable study in recent years, beginning with the work of Erdős, Mauduit, and Sárközy, who studied the distribution of residues in $\mathcal{K}(S, b)$ modulo small numbers [3] and proved the existence of integers in $\mathcal{K}(S, b)$ with many small prime factors [4]. Notable recent work includes Maynard's proof [7] that the sets $\mathcal{K}(S, b)$ contain infinitely many primes whenever $b - |S|$ is at most $b^{23/80}$, provided b is sufficiently large.

In this article we consider the additive structure of proper Kempner sets. In particular, we consider the following extremal question: *What is the length of the longest arithmetic progression in a proper Kempner set with a fixed given base?* Our methods will be combinatorial, rather than analytic (as in Maynard's work, [7]).

A well-known conjecture of Erdős states that any set of positive integers with a divergent harmonic sum contains arithmetic progressions of arbitrary (finite) length. Since proper Kempner sets have convergent harmonic sums, this might suggest that the lengths of arithmetic progressions in a given proper Kempner set are uniformly bounded.

This is indeed the case. Let us say that a set $T \subset \mathbb{Z}$ is k -free if T contains no arithmetic progression of length k . By a simple argument, given in Proposition 3, one may show that the proper Kempner set $\mathcal{K}(S, b)$ is $(b^2 - b + 1)$ -free for any $b \geq 2$.

The main purpose of this article is to understand how close this trivial upper bound is to the truth.

In our main theorem, we improve this bound for all $b \geq 2$, obtaining a tight result that expresses the length of the longest arithmetic progression in $\mathcal{K}(S, b)$ in terms of the prime factorization of b . To state this theorem, we need to introduce some arithmetic functions. If n and b are natural numbers, let $\rho(n)$ denote the square-free radical of n (i.e., the product of all distinct primes dividing n), and let $\beta(b)$ denote the largest integer less than b such that $\rho(\beta(b))|b$. For example, $\beta(10) = 8$, and $\beta(p^k) = p^{k-1}$ for any prime power p^k . In other words, $\beta(b)$ is the greatest integer less than b that divides some power of b . Finally, let $\ell(b)$ be the length of the longest arithmetic progression contained in some proper Kempner set of base b .

Our main theorem gives an exact evaluation of $\ell(b)$.

Theorem 1. *For all $b \geq 2$, one has $\ell(b) = (b - 1)\beta(b)$.*

For example, $\ell(10) = 72$. One particular set that achieves this bound is Kempner's original set, $\mathcal{K}([0, 8], 10)$, which contains the 72-term arithmetic progression $\{0, 125, 250, 375, \dots, 8875\}$.

The arithmetic functions $\beta(b)$ and $\ell(b)$ are of independent interest, but do not appear to have been considered seriously before.¹ We establish average order results for $\beta(b)$ which show that, for most b , the trivial upper bound on $\ell(b)$ from Proposition 3 is asymptotically correct.

Theorem 2. *There is a set of integers $A \subset \mathbb{Z}$ with natural density 1, i.e., with*

$$\lim_{N \rightarrow \infty} \frac{1}{N} |A \cap [1, N]| = 1,$$

such that $\ell(b) \sim b^2$ as $b \rightarrow \infty$ in A .

Notation: For $x \in \mathbb{R}$, let $\{x\}$ denote the fractional part of x and let $\lfloor x \rfloor$ denote the greatest integer that is at most x . For a natural number n , we let $[n]$ denote the set of integers $\{1, \dots, n\}$. As mentioned previously, for two integers x and y we use $[x, y]$ to denote the set $\{n \in \mathbb{Z} : x \leq n \leq y\}$. We use the notation $\log_q p$ to denote the logarithm of p to base q (as opposed to any iterations of logarithms). The expression $\log x$ will be interpreted as the natural log.

2. PROGRESSIONS OF MAXIMAL LENGTH IN KEMPNER SETS. In this section we give our proof of Theorem 1, which is an exact evaluation of $\ell(b)$ and the main result of this article. This will be done in two parts: providing a constructive lower bound and giving a proof that this lower bound is sharp. Before that, as promised, we give a simple proof that the function $\ell(b)$ is at least well-defined, i.e., that Kempner sets do not contain arbitrarily long arithmetic progressions.

Proposition 3. *For all $b \geq 2$, we have $\ell(b) \leq (b - 1)b$.*

Proof. Suppose that $A \subset \mathcal{K}(S, b)$ is a finite arithmetic progression of $|A|$ terms with common difference Δ . Choose $k \geq 0$ such that $b^k \leq \Delta < b^{k+1}$. If I denotes the shortest interval of integers containing A , then $|I| = (|A| - 1)\Delta + 1$; hence $|A| = 1 + (|I| - 1)/\Delta$.

¹The sequence $\beta(b)$ is entry A079277 on the Online Encyclopedia of Integer Sequences; see oeis.org/A079277

Suppose that A excludes the digit d . Then certainly A is a subset of the union of the following disjoint intervals:

$$[0, db^{k+1} - 1] \quad (1)$$

and the intervals

$$[b^{k+2}m + (d+1)b^{k+1}, b^{k+2}(m+1) + db^{k+1} - 1] \quad (2)$$

as m ranges over $\mathbb{Z}_{\geq 0}$. Indeed, all the natural numbers that are not contained in this union have the digit d in their base- b expansion (in the place corresponding to b^{k+1}).

We now prove that A must be contained within just one of these intervals. If $|A| = 1$, then the statement is trivial, so we may assume that $|A| \geq 2$. Let x denote the least element of A and suppose that x is an element of the interval J (where J is of the form (1) or (2)). Then it follows by induction on n that $x + n\Delta \in J$ for all integers n in the range $0 \leq n \leq |A| - 1$. Indeed, the base case $n = 0$ holds by definition, so let us choose n in the range $1 \leq n \leq |A| - 1$ and assume the induction hypothesis that $x + (n-1)\Delta \in J$. Since $n \leq |A| - 1$ we have $x + n\Delta \in A$, and so certainly $x + n\Delta$ is a member of one of the intervals of the form (1) or (2). Let us call this interval J' . If $J' \neq J$, it follows that if $y' \in J'$ and $y \in J$ then $|y' - y| > b^{k+1}$. Since $x + (n-1)\Delta \in J$ by the inductive hypothesis, if $J' \neq J$ we would have

$$\Delta = |(x + n\Delta) - (x + (n-1)\Delta)| > b^{k+1},$$

which contradicts our assumption that $\Delta < b^{k+1}$. Therefore $x + n\Delta \in J$ too, and the inductive step is complete.

Thus $A \subset J$ and so $|I| \leq |J|$. This gives $|I| \leq b^{k+2} - b^{k+1}$, which yields

$$|A| \leq 1 + \frac{b^{k+2} - b^{k+1} - 1}{\Delta} < 1 + \frac{b^{k+2} - b^{k+1}}{b^k} \leq b^2 - b + 1;$$

hence $|A| \leq b^2 - b$ as claimed. \blacksquare

The bound in the previous proposition is simple and—as a consequence—occasionally weak. In particular, it neglects the potentially compounding effects of digit exclusion at different orders of magnitude, and the arithmetic properties of orbits in the group $\mathbb{Z}/b\mathbb{Z}$. This structure can affect the bounds dramatically, as seen most clearly in the case when the base b is prime.

Proposition 4. *Let p be prime. Then $\ell(p) \leq p - 1$.*

Proof. Suppose that $\mathcal{K}(S, p)$ is a proper Kempner set and contains the progression $A = \{k + j\Delta : j \in [p]\}$ with $\Delta \neq 0$. Since $|S| \leq p - 1$, $|A|$ cannot cover all the residue classes modulo p , and so by the pigeonhole principle there exist distinct $i, j \in [p]$ with $k + j\Delta \equiv k + i\Delta \pmod{p}$; hence $p \mid \Delta$ (since p is prime). By deleting the rightmost digits of the elements of A , we obtain a new progression in $\mathcal{K}(S, p)$ with common difference Δ/p ; in particular, the progression

$$\left\{ \left\lfloor \frac{k}{p} \right\rfloor + j \frac{\Delta}{p} : j \in [p] \right\}.$$

The new common difference is strictly smaller, and we obtain a contradiction by infinite descent. \blacksquare

With a little more bookkeeping this proof generalizes to prime powers, and implies that $\ell(p^k) \leq p^{k-1}(p^k - 1)$. So certainly $\ell(b)$ is not asymptotic to b^2 as b ranges over all integers; some restriction in Theorem 2 is required.

We now begin the proof of Theorem 1. Searching for long progressions in $\mathcal{K}([0, 8], 10)$, one might happen across the example noted earlier, namely the first 71 multiples of 125, which—together with 0—form an arithmetic progression of length 72, none of whose members contain the digit 9. This example succeeds due to properties of the prime factorization of $1000/125$ in relation to the base 10. These properties generalize, and one may use this to construct long digit-excluding arithmetic progressions in arbitrary bases.

Proposition 5. *For all $b \geq 2$, the Kempner set $\mathcal{K}([0, b-2], b)$ contains an arithmetic progression of length $(b-1)\beta(b)$. Hence $\ell(b) \geq (b-1)\beta(b)$.*

Proof. Let $K \geq 1$ be the smallest natural number such that $\beta(b) \mid b^K$. We claim that all the members of the arithmetic progression

$$A = \frac{b^K}{\beta(b)}[0, (b-1)\beta(b) - 1]$$

exclude the digit $b-1$ from their base- b expansions. To see this, let k satisfy $0 \leq k \leq K-1$. Then $\gcd(b^{k+1}\beta(b), b^K) \geq b^{k+1} > b^k\beta(b)$, which implies that $\gcd(b^{k+1}, b^K/\beta(b)) > b^k$ (by dividing through by $\beta(b)$). In particular, for all integers x and y , either

$$\left| x \frac{b^K}{\beta(b)} - yb^{k+1} \right| > b^k \quad \text{or} \quad x \frac{b^K}{\beta(b)} = yb^{k+1}. \quad (3)$$

This observation implies that none of the K rightmost digits of any integer of the form $xb^K/\beta(b)$ can be equal to $b-1$. Indeed, in base b , the b^k digit of $xb^K/\beta(b)$ is the unique integer d in the range $0 \leq d \leq b-1$ such that

$$\left\{ \frac{xb^K/\beta(b)}{b^{k+1}} \right\} \in \left[\frac{d}{b}, \frac{d+1}{b} \right).$$

Yet (3) implies that $\left\{ \frac{xb^K}{b^{k+1}\beta(b)} \right\} \in \{0\} \cup (\frac{1}{b}, \frac{b-1}{b})$ for each $0 \leq k \leq K-1$. Since this is disjoint from $[\frac{b-1}{b}, 1)$, we conclude that none of the K rightmost digits of any integer of the form $xb^K/\beta(b)$ can be equal to $b-1$.

We now fix $x \in [0, (b-1)\beta(b) - 1]$ and consider the leftmost digits of $xb^K/\beta(b)$. Certainly $xb^K/\beta(b) < (b-1)b^K$. From this upper bound we see that the b^K digit of $xb^K/\beta(b)$ lies in $[0, b-2]$ and that the digits associated to larger powers of b are all 0. Combining this with our previous observations, we conclude that $xb^K/\beta(b)$ omits the digit $(b-1)$ for all $x \in [0, (b-1)\beta(b) - 1]$, so $A \subset \mathcal{K}([0, b-2], b)$ as claimed. Since $|A| = (b-1)\beta(b)$, we have $\ell(b) \geq (b-1)\beta(b)$. ■

We now proceed with the second half of our evaluation of $\ell(b)$, the verification that this lower bound is exact. This requires a more technical argument.

Proposition 6. *For all $b \geq 2$, we have $\ell(b) \leq (b-1)\beta(b)$.*

Proof. Without loss of generality, let $S \subset [0, b-1]$ be any set of $b-1$ digits (containing 0), and let $A = \{x + j\Delta : j \in [0, \ell(b) - 1]\}$ be an arithmetic progression in $\mathcal{K}(S, b)$ of maximal length, in which $\Delta > 0$ is taken minimally over all arithmetic progressions of length $\ell(b)$.

Let $\Delta = d_K b^K + \cdots + d_1 b + d_0$ denote the base b expansion of Δ , where K is chosen such that $d_K \neq 0$. For notational convenience, let $\Delta_k := d_k b^k + \cdots + d_1 b + d_0$ for each $k \geq 0$. (Note that $\Delta_k = \Delta$ for $k \geq K$.) We may assume without loss of generality that $d_0 \neq 0$, else by removing the rightmost digit from all elements of A one constructs an arithmetic progression contained in $\mathcal{K}(S, b)$ of common difference Δ/b , contradicting our minimality assumption on Δ . (This is the same device as we used in the proof of Proposition 4.)

Our proof of Proposition 6 rests on the following claim, whose peculiar statement arises naturally from an inductive argument.

Claim 7. *Consider the following statements:*

C1: $\ell(b) \leq (b-1)\beta(b)$;

C2(k): *there exist coprime integers $\lambda_k, \mu_k \in [1, b-1]$ satisfying $\lambda_k \Delta_k = \mu_k b^{k+1}$.*

Then either C1 holds or C2(k) holds for all $k \geq 0$.

This claim immediately settles the theorem, since the statement C2(k) cannot possibly hold for all $k \geq 0$. Indeed, we have $\lambda_k \Delta_k < b\Delta$, while $\mu_k b^{k+1}$ grows in k without bound. ■

Proof of Claim. We prove this claim by induction, showing that for every $k \geq 0$, either C1 holds or C2(k') holds for all $k' \leq k$. For the base case $k = 0$, note that $\Delta_k = d_0$. If $(d_0, b) = 1$, then d_0 generates the additive group $\mathbb{Z}/b\mathbb{Z}$ and the elements $\{x + j\Delta : j \in [0, b-1]\}$ have b distinct units digits. Thus $\ell(b) \leq (b-1) \leq (b-1)\beta(b)$, so C1 holds.

Otherwise, $(d_0, b) > 1$, which implies that there exists $\lambda \in [1, b-1]$ for which $\lambda d_0 \equiv 0 \pmod{b}$. Thus $\lambda d_0 = \mu b$ for some μ , and we may assume that $(\lambda, \mu) = 1$ by dividing through by common factors. This concludes the base case.

Proceeding to the inductive step, let $k \geq 1$ and assume that the inductive hypothesis C2(k') holds for all smaller k' . In particular, $\Delta_{k-1} = (\mu_{k-1}/\lambda_{k-1})b^k$ for some coprime integers $\lambda_{k-1}, \mu_{k-1} \in [1, b-1]$, and hence $\Delta_k = d_k b^k + (\mu_{k-1}/\lambda_{k-1})b^k$.

Let λ_k denote the order of Δ_k/b^{k+1} in the additive group \mathbb{R}/\mathbb{Z} , and let μ_k denote the integer $\lambda_k(\Delta_k/b^{k+1})$. We see that $(\lambda_k, \mu_k) = 1$, as one could divide through by any common factors of λ_k and μ_k to contradict the fact that λ_k is the order of Δ_k/b^{k+1} in \mathbb{R}/\mathbb{Z} . Now, if $\lambda_k < b$, then $\mu_k < b$ as well, since $\Delta_k/b^{k+1} < 1$ for any k . In this case, λ_k and μ_k satisfy the conditions listed in C2(k). Therefore C2(k') holds for all $k' \leq k$.

It remains to address the case $\lambda_k \geq b$. By usual facts about finite subgroups of \mathbb{R}/\mathbb{Z} , we note that the orbit of Δ_k/b^{k+1} in \mathbb{R}/\mathbb{Z} is exactly the set of fractions with denominator dividing λ_k . In particular, the set of values

$$T = \left\{ \frac{x}{b^{k+1}} + \frac{\Delta_k j}{b^{k+1}} \pmod{1} : j \in [0, \lambda_k - 1] \right\}$$

are equally spaced, with gaps of size $1/\lambda_k$. Since $\lambda_k \geq b$, for any integer $d \in [0, b-1]$ at least one member of T lies in the half-open interval $[\frac{d}{b}, \frac{d+1}{b})$. In other words, at least one member of the progression $x + \Delta_k[0, \lambda_k - 1]$ has b^k digit equal to d .

This information immediately implies that $x + \Delta[0, \lambda_k - 1]$ is not contained in any proper Kempner set $\mathcal{K}(S, b)$, and hence $\ell(b) \leq \lambda_k - 1$. However, more can be said with a slight refinement to our analysis. Equal spacing implies that at least $\lfloor \lambda_k/b \rfloor$ members of T lie in the interval $[\frac{d}{b}, \frac{d+1}{b})$. We are left with the stronger bound $\ell(b) \leq \lambda_k - \lfloor \lambda_k/b \rfloor$.

We now establish an upper bound on the function $\lambda_k - \lfloor \lambda_k/b \rfloor$, given the known constraints on λ_k . For starters, the inductive hypothesis implies that $\lambda_{k-1} \mid \mu_{k-1}b^k$, hence $\lambda_{k-1} \mid b^k$ (since λ_{k-1} and μ_{k-1} are coprime). Since $\lambda_{k-1} < b$ and λ_{k-1} divides a power of b , this implies that $\lambda_{k-1} \leq \beta(b)$. Second, the inductive hypothesis allows us to write

$$\frac{\Delta_k}{b^{k+1}} = \frac{d_k \lambda_{k-1} + \mu_{k-1}}{b \lambda_{k-1}},$$

which implies that $b \lambda_{k-1} (\Delta_k/b^{k+1}) \equiv 0 \pmod{1}$. This implies that $b \lambda_{k-1}$ is a multiple of the order of $(\Delta_k/b^{k+1}) \pmod{1}$, i.e., $\lambda_k \mid b \lambda_{k-1}$. We conclude that $\lambda_k \leq b \lambda_{k-1} \leq b \beta(b)$.

The function $\lambda \mapsto \lambda - \lfloor \lambda/b \rfloor$ is nondecreasing as λ increases over integers, hence

$$\ell(b) \leq \lambda_k - \left\lfloor \frac{\lambda_k}{b} \right\rfloor \leq b \beta(b) - \left\lfloor \frac{b \beta(b)}{b} \right\rfloor = b \beta(b) - \beta(b) = (b-1) \beta(b),$$

which implies that C1 holds. This completes the inductive step, and so completes the proof of Theorem 1. ■

3. ASYMPTOTIC ANALYSIS. In this section we analyze the function $\beta(b)$, with the ultimate goal of proving Theorem 2. We begin with the following simple observation.

Proposition 8. *We have*

$$\liminf_{n \rightarrow \infty} \frac{\beta(n)}{n} = 0 \quad \text{and} \quad \limsup_{n \rightarrow \infty} \frac{\beta(n)}{n} = 1.$$

Proof. The first claim follows from the observation that $\beta(p) = 1$ for all primes p . For the second, we note that $\beta(2^k + 2) = 2^k$ for all $k > 1$. ■

It is clear from this proposition that the behavior of $\beta(n)$ is erratic as n varies. However, its calculation may be understood as a certain integer programming problem, as illustrated by the following example.

Example 9. In this example, we calculate $\beta(24)$ using techniques from mixed integer programming. We may write $\beta(24) = 2^a \cdot 3^b$, with $a, b \in \mathbb{N}$. It follows that $a \log 2 + b \log 3 < \log 24$, and (a, b) may be visualized as a lattice point in Figure 1. The equation of the line is $f(x) = \log_3 24 - x \log_3 2$.

Let us restrict our attention to the set S of lattice points of the form (a, b) , in which b is taken maximally for fixed a . If $(a, b) \in S$, then the vertical distance from (a, b) to the diagonal in Figure 1 is then given by $\{\log_3 24 - a \log_3 2\}$. We also note that $a \log 2 + b \log 3$ is maximized among the lattice points below the line when $(a, b) \in S$ and $\{\log_3 24 - a \log_3 2\}$ is minimized (as a function of a). In our example, minimization occurs at $(a, b) = (1, 2)$, and so we obtain $\beta(24) = 2^1 \cdot 3^2 = 18$.

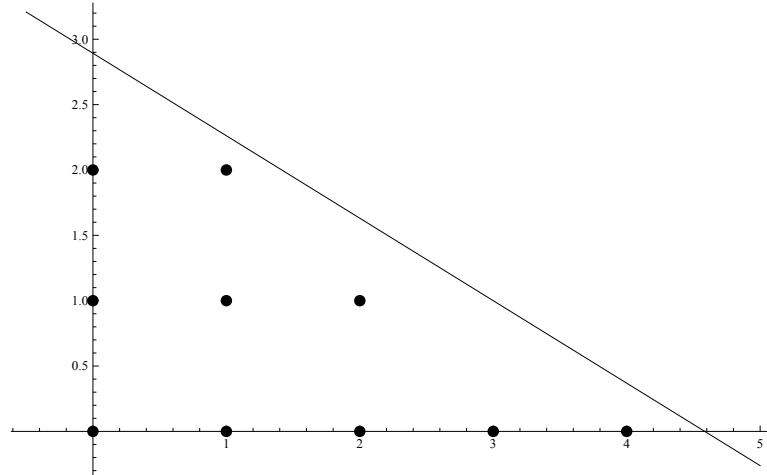


Figure 1. Lattice points (a, b) corresponding to $\beta(24) = 2^a \cdot 3^b$.

The technique of Example 9 generalizes easily: if n has k prime divisors p_1, \dots, p_k , we may associate to n a set of lattice points in \mathbb{Z}^k , namely

$$\{(a_1, \dots, a_k) \in \mathbb{Z}_{\geq 0}^k : a_1 \log p_1 + \dots + a_k \log p_k < \log n\}.$$

The lattice point (a_1, \dots, a_k) that minimizes distance to the the hyperplane

$$x_1 \log p_1 + \dots + x_k \log p_k = \log n$$

determines $\beta(n)$ by the formula $\beta(n) = \prod_{i=1}^k p_i^{a_i}$.

Combining this idea with well-known equidistribution results gives the following.

Lemma 10. *We have $\beta(n) \sim n$ as $n \rightarrow \infty$ within $N\mathbb{Z}$ if and only if N is not a prime power.*

Proof. If $N = p^k$ is a prime power, then $N\mathbb{Z}$ contains the subsequence $\{p^{kn}\}_{n \geq 1}$. Since $\beta(p^{kn}) = p^{kn-1}$, we cannot have $\beta(n) \sim n$ within $N\mathbb{Z}$.

Otherwise, let p and q be distinct primes dividing N , and fix a positive constant ε . As $\log_q p$ is irrational, the sequence $(\{u_n\})_{n=0}^\infty$ given by $u_n := n \cdot \log_q p$ is equidistributed mod 1 (by the equidistribution theorem: see [5, Proposition 21.1], say).

To spell it out, this theorem states that for every pair of real numbers λ_1, λ_2 that satisfy $0 \leq \lambda_1 \leq \lambda_2 \leq 1$,

$$\frac{1}{N} |\{n \leq N : \lambda_1 \leq \{u_n\} \leq \lambda_2\}| \rightarrow \lambda_2 - \lambda_1$$

as $N \rightarrow \infty$. In particular, by splitting $[0, 1]$ into disjoint intervals of length $\varepsilon/2$, we see that there exists a positive parameter L_ε such that $l > L_\varepsilon$ implies that the sequence $(\{u_n\})_{n=0}^l$ contains at least one element in each interval mod 1 of length ε .

Now let m be a natural number and let $l = \lfloor \log_p m \rfloor$. From the above remarks, there exists a positive parameter M_ε such that, for each $m > M_\varepsilon$, the shifted sequence $(\{\log_q m - u_n\})_{n=0}^l$ contains some element in the interval $(0, \varepsilon)$. In other words there exists n_0 at most l (but dependent on l) such that

$$0 < \{\log_q m - n_0 \cdot \log_q p\} < \varepsilon.$$

Also note that $\log_q m - n_0 \cdot \log_q p$ is positive.

Now, assume $pq \mid m$ and consider $(a, b) := (n_0, \lfloor \log_q m - n_0 \cdot \log_q p \rfloor)$. We have $\beta(m) \geq p^a \cdot q^b$ by construction. So

$$\beta(m) \geq p^a \cdot q^b = q^{\log_q m - \{\log_q m - n_0 \cdot \log_q p\}} > q^{\log_q m - \varepsilon} = m \cdot q^{-\varepsilon}.$$

Thus $q^{-\varepsilon} < \beta(m)/m < 1$, for all m satisfying $m > M_\varepsilon$ and $pq \mid m$. Since ε was arbitrary, and q fixed, it follows that $\beta(m) \sim m$ within $pq\mathbb{Z}$, and hence within $N\mathbb{Z}$. ■

By considering $N = 6$, for example, we obtain a set of density $1/6$ (namely, $6\mathbb{Z}$) on which $\ell(b) \sim b^2$ as b tends to infinity within that set. Any finite union of such sets $N_i\mathbb{Z}$, where N_i has two distinct prime factors p_i and q_i , will also have this property, and one may show with relative ease that such a union may be arranged to have natural density arbitrarily close to 1.

However, by quantifying estimates made in the previous lemma, we can do slightly better and show the existence of a set with the desired property that has density 1, thereby proving Theorem 2.

Proof of Theorem 2. Let $f(N)$ be a nondecreasing function that satisfies $f(N) \rightarrow \infty$ as $N \rightarrow \infty$ (to be further specified later). For integers $j \geq 0$, let D_j denote the set of $n \in (2^{j-1}, 2^j]$ such that n has at least two distinct prime factors $p, q \leq f(2^{j-1})$. Let

$$D := \bigcup_{j \geq 0} D_j.$$

The set D is our candidate set for use in Theorem 2.

Lemma 11. *If f grows slowly enough, the set D has natural density 1.*

Proof. We begin by fixing $j \geq 0$ and bounding the size of D_j from below. For convenience, we write N for 2^{j-1} .

To produce this lower bound, we find an upper bound for $(N, 2N] \setminus D_j$. Indeed, if $n \in (N, 2N] \setminus D_j$ then either n has no prime factor less than $f(N)$ or it has exactly one such prime factor. To bound the number of n of the first type, by a standard application of a small sieve (e.g., the Selberg sieve, in particular [9, Theorem 9.3.10]) one may show that the number of such n is

$$O\left(N \prod_{p < f(N)} \left(1 - \frac{1}{p}\right)\right), \quad (4)$$

where the product runs over primes p , provided $f(N)$ grows slowly enough ($f(N) \leq N^{1/10}$ will certainly suffice). One recalls Mertens' formula (sometimes called "Mertens' second theorem" or "Mertens' third theorem"):

$$\prod_{p < x} \left(1 - \frac{1}{p}\right) \sim \frac{e^{-\gamma}}{\log x}$$

as $x \rightarrow \infty$, where $\gamma = 0.577 \dots$ is the Euler–Mascheroni constant (see [11, Theorem 1.12]). We conclude that the quantity (4) is $O(N/\log f(N))$.

Our next step is to bound the number of $n \in (N, 2N]$ with exactly one prime factor less than $f(N)$. By using the map $n \mapsto pm$, this quantity may be bounded above by the number of pairs (p, m) for which $p < f(N)$, $m \in (N/p, 2N/p]$, and m has no

prime factor less than $f(N)$. By using the sieve bound as above, the number of such pairs (p, m) is

$$O\left(\sum_{p < f(N)} \frac{N}{p} \prod_{\substack{q < f(N) \\ q \neq p}} \left(1 - \frac{1}{q}\right)\right).$$

This quantity is $O(N \log \log f(N) / \log f(N))$, using Mertens' formula as above together with the related result

$$\sum_{p < x} \frac{1}{p} = O(\log \log x)$$

(see [11, Theorem 1.10]). We conclude by exclusion that

$$|D_j| = N \left(1 - O\left(\frac{\log \log f(N)}{\log f(N)}\right)\right).$$

This already establishes that D has full upper Banach density. To show that D has natural density 1, we fix $\varepsilon > 0$ and note that, since $f(N) \rightarrow \infty$ as $j \rightarrow \infty$, there exists $j_0(\varepsilon)$ such that $|D_j| \geq 2^{j-1}(1 - \varepsilon)$ for all $j \geq j_0(\varepsilon)$. In particular,

$$\begin{aligned} \sum_{\substack{n \leq X \\ n \in D}} 1 &\geq \sum_{j_0(\varepsilon) \leq j \leq \lceil \log_2 X \rceil} |D_j| - \sum_{X < n \leq 2^{\lceil \log_2 X \rceil}} 1. \\ &\geq (1 - \varepsilon) \left(2^{\lceil \log_2 X \rceil} - 2^{j_0(\varepsilon)-1}\right) + X - 2^{\lceil \log_2 X \rceil}. \end{aligned}$$

Simplifying, we see that

$$\liminf_{X \rightarrow \infty} \frac{|D \cap [1, X]|}{X} \geq \liminf_{X \rightarrow \infty} \frac{X - \varepsilon 2^{\lceil \log_2 X \rceil} - 2^{j_0(\varepsilon)}}{X} \geq 1 - 2\varepsilon,$$

which implies that D has natural density 1, since ε was arbitrary. ■

Second, we prove that $\beta(n)$ is asymptotically large within D .

Lemma 12. *If f grows slowly enough, then $\beta(n) \sim n$ as $n \rightarrow \infty$ within D .*

Proof. Our proof presents a more quantitative adaptation of the argument used in Lemma 10. Let $\varepsilon > 0$, and fix $n \in D$. By the definition of D , there exist distinct primes $p, q < f(n)$ for which $p, q \mid n$. We will show that, provided n is large enough in terms of ε , there exist nonnegative integers a and b for which

$$e^\varepsilon \geq \frac{n}{p^a q^b} > 1.$$

Since $p^a q^b \leq \beta(n) < n$, and ε is arbitrary, this will complete the proof.

Taking logarithms, it suffices to find nonnegative integers a and b for which

$$\frac{\varepsilon}{\log q} \geq \log_q n - a \log_q p - b > 0.$$

Setting $L = \lfloor \log_p n \rfloor$, it will be enough to prove that the sequence of fractional parts $\{a \log_q p : a \in [1, L]\}$ contains an element in every interval modulo 1 of length $\varepsilon / \log q$. Since $p, q \leq f(n)$, we reduce our theorem to the following claim:

Claim 13. *Let $L' = \lfloor \log n / \log f(n) \rfloor$. Then $S = \{a \log_q p : a \in [1, L']\}$ contains an element in every interval modulo 1 of length $\varepsilon / \log f(n)$, provided $f(n)$ grows slowly enough.*

The proof of this claim follows from the Erdős–Turán inequality (see [8, Corollary 1.1]). Indeed, for any interval I modulo 1 of length $\varepsilon / \log f(n)$, we have

$$\left| |S \cap I| - \frac{\varepsilon L'}{\log f(n)} \right| \ll \frac{L'}{K+1} + \sum_{k \leq K} \frac{1}{k} \left| \sum_{a=1}^{L'} e^{2\pi i a k \log_q p} \right| \quad (5)$$

for any integer $K \geq 1$. It suffices to show that we may choose a K such that the right-hand side in (5) is $o(L' / \log f(n))$ as $n \rightarrow \infty$.

Choosing $K = \lfloor \log^2 f(n) \rfloor$ ensures that $L' / (K+1) = o(L' / \log f(n))$. As for the second term in (5), bounding the sum over a as a geometric series gives

$$\sum_{k \leq K} \frac{1}{k} \left| \sum_{a=1}^{L'} e^{2\pi i a k \log_q p} \right| \leq G(K, p, q)$$

for some function G that is independent of L' . We may assume without loss of generality that G is nondecreasing in each variable. Then

$$G(K, p, q) \ll G(\log^2 f(n), f(n), f(n)),$$

so it suffices to show that

$$G(\log^2 f(n), f(n), f(n)) = o\left(\frac{L'}{\log f(n)}\right). \quad (6)$$

Recalling the definition of L' , this is equivalent to showing

$$G(\log^2 f(n), f(n), f(n)) \cdot \log^2 f(n) = o(\log n).$$

Yet G is simply some absolute function, so if f grows slowly enough then (6) will hold. (If one so wished, one could quantify this growth condition using Baker's result [2] on linear forms of logarithms of primes). This proves the claim, and hence the lemma. ■

Combining Lemma 12 with Theorem 1 yields Theorem 2. ■

ACKNOWLEDGMENTS. The authors would like to thank the Mathematical Sciences Research Institute, Berkeley, California, for providing excellent working conditions during the Analytic Number Theory Semester of 2017, when the majority of the work involved in this article was carried out. During this time the first named author was partially supported by EPSRC Grant EP/M50659X/1. The authors would also like to thank two anonymous referees and the journal editor for corrections to earlier drafts.

REFERENCES

- [1] Baillie, R. (1979). Sums of reciprocals of integers missing a given digit. *Amer. Math. Monthly*. 86(5): 372–374. doi.org/10.2307/2321096
- [2] Baker, A. (1968). Linear forms in the logarithms of algebraic numbers. IV. *Mathematika*. 15: 204–216. doi.org/10.1112/S0025579300002588
- [3] Erdős, P., Mauduit, C., Sárközy, A. (1998). On arithmetic properties of integers with missing digits I. Distribution in residue classes. *J. Number Theory*. 70(2): 99–120. doi.org/10.1006/jnth.1998.2229
- [4] Erdős, P., Mauduit, C., Sárközy, A. (1999). On arithmetic properties of integers with missing digits II. Prime factors. *Discrete Math*. 200(1-3): 149–164. doi.org/10.1016/S0012-365X(98)00331-8
- [5] Iwaniec, H., Kowalski, E. (2004). *Analytic Number Theory*, American Mathematical Society Colloquium Publications, vol.53. Providence, RI: American Mathematical Society.
- [6] Kempner, A. J. (1914). A curious convergent series. *Amer. Math. Monthly*. 21(2): 48–50. doi.org/10.2307/2972074
- [7] Maynard, J. (2016). Primes with restricted digits (preprint). Available at arxiv.org/abs/1604.01041
- [8] Montgomery, H. L. (1994). *Ten Lectures on the Interface between Analytic Number Theory and Harmonic Analysis*, CBMS Regional Conference Series in Mathematics, vol. 84. Providence, RI: American Mathematical Society.
- [9] Murty, M. R. (2008). *Problems in Analytic Number Theory*, Graduate Texts in Mathematics, vol. 206. New York, NY: Springer.
- [10] Schmelzer, T., Baillie, R. (2008). Summing a curious, slowly convergent series. *Amer. Math. Monthly*. 115(6): 525–540. doi.org/10.1080/00029890.2008.11920559
- [11] Tenenbaum, G. (2015). *Introduction to Analytic and Probabilistic Number Theory*, 3rd ed. Graduate Studies in Mathematics, vol. 163. Providence, RI: American Mathematical Society.

ALED WALKER is a Junior Research Fellow at Trinity College Cambridge, having recently completed his DPhil at Magdalen College Oxford under the supervision of Ben Green. His research interests include analytic number theory, additive combinatorics, and the local statistics of point sequences.

Trinity College, Cambridge, CB2 1TQ, United Kingdom
aw530@cam.ac.uk

ALEXANDER WALKER is a postdoctoral fellow at Rutgers University and a recent graduate of the doctoral program at Brown University, where he earned his degree under the direction of Jeffrey Hoffstein. His research interests include analytic number theory and automorphic forms.

Department of Mathematics, Rutgers University, Piscataway, NJ 08854
alexander.walker@rutgers.edu